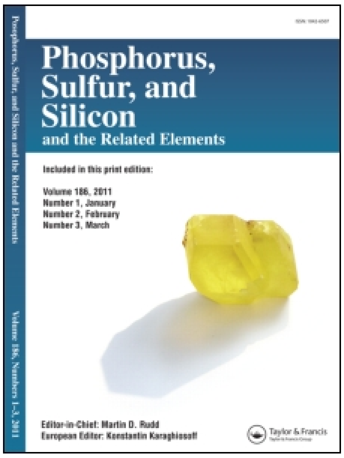
This article was downloaded by:
On: *28 January 2011*
Access details: *Access Details: Free Access*
Publisher *Taylor & Francis*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Phosphorus, Sulfur, and Silicon and the Related Elements

Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713618290

Comparison of Some Computational Methods for Geometry Optimisation of Phosphorus Acid Derivatives

Susannah Sigurdsson[a]; Roger Strömberg[a]
[a] Division of Organic and Bioorganic Chemistry, Scheele Laboratory, Karolinska Institutet, Stockholm, Sweden

Online publication date: 27 October 2010

To cite this Article Sigurdsson, Susannah and Strömberg, Roger(2002) 'Comparison of Some Computational Methods for Geometry Optimisation of Phosphorus Acid Derivatives', Phosphorus, Sulfur, and Silicon and the Related Elements, 177: 11, 2711 — 2724

To link to this Article: DOI: 10.1080/10426500214562

URL: http://dx.doi.org/10.1080/10426500214562

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Phosphorus, Sulfur and Silicon, 2002, Vol. 177:2711–2724
Copyright © 2002 Taylor & Francis
1042-6507/02 $12.00 + .00
DOI: 10.1080/10426500290110694

COMPARISON OF SOME COMPUTATIONAL METHODS FOR GEOMETRY OPTIMISATION OF PHOSPHORUS ACID DERIVATIVES

Susannah Sigurdsson and Roger Strömberg
Division of Organic and Bioorganic Chemistry,
Scheele Laboratory, Karolinska Institutet, Stockholm, Sweden

(Received February 8, 2002; accepted April 12, 2002)

Several economical methods for geometry optimization, that should be applicable to larger molecules, have been evaluated for 19 phosphorus acid derivatives. MP2/cc-pVDZ geometry optimizations are used as reference points and the geometries obtained from the other methods are evaluated with respect to deviations in bond lengths and angles, from the reference geometries. The geometry optimization methods are also compared to the much used B3LYP/6-31G(d) method. Single point energies obtained by subsequent EDF1/6-31+G(d) or B3LYP/6-31+G(d,p) calculations on the respective equilibrium geometries are also reported relative to the energies obtained from the reference geometries.

The geometries from HF/MIDI! optimizations were closer to those of the references than the geometries of the HF/3-21G(d), HF/6-31G(d), and B3LYP/MIDI! optimizations. The EDF1/6-31+G(d) or B3LYP/6-31+G(d,p) single point energies obtained from the HF/3-21G(d), HF/6-31G(d), and B3LYP/MIDI! geometries gave a mean absolute deviation (MAD) from that of the reference geometries of 1.4–3.9 kcal mol^{-1}*. The HF/MIDI! geometries, however, gave EDF1/6-31+G(d) and B3LYP/6-31+G(d,p) energies with a MAD of only about 0.5 and 0.55 kcal mol*$^{-1}$ *respectively from the energies obtained with the reference geometries. Thus, use of HF/MIDI! for geometry optimization of phosphorus acids is a method that gives geometries of near-MP2 quality, resulting in a fair accuracy of energies in subsequent single point calculations, at a much lower computational cost other methods that give similar accuracies.*

Keywords: Ab initio; EDF1; Hartree Fock; MIDI!; MP2

We thank the Swedish Natural Science Research Council and the Swedish Research Council for Engineering Sciences for financial support.

Address correspondence to Roger Strömberg, Division of Organic and Bioorganic Chemistry, MBB, Scheele Laboratory, Karolinska Institutet, S-171-77 Stockholm, Sweden. E-mail: Roger.Stromberg@mbb.ki.se

INTRODUCTION

With the fast development in processing power of personal computers and workstations the number of chemists using quantum mechanical calculations as a tool in their research has increased dramatically. However, the level of accuracy that can be afforded for nonspecialists is usually quite limited since the smallest desirable size of a system often contains more than 10 heavy atoms. In addition, many systems for biologically related problems are even larger, and often contain third row atoms like Mg, P, and S. It is therefore of great concern to evaluate methods that can be applied to chemical problems with reasonable accuracy, and at a computational cost that is affordable by scientists with limited computing facilities who wish to use quantum chemical calculations as a complement to their experimental research. The time-limiting step for most model chemistries is usually the geometry optimization.

Recently Hartree-Fock (HF) calculation with the MIDI![1] basis set has been forwarded as a cost-efficient method for achieving geometries of reasonably good quality (near MP2-quality). The geometry deviations from comparison with MP2/cc-pVDZ calculations were presented in the form of a composite of bond length and bond angle deviations and the impact on subsequent single point energy calculations was not reported. As opposed to most basis functions the MIDI! set was developed in order to predict accurate geometries rather than energies. Calculation strategies often contain initial optimization at the Hartree-Fock level with small basis sets, such as 3-21G(d) or 6-31G(d). This is generally followed by higher level methods for obtaining more accurate energies, most successfully in compound methods such as CBS-4.[2] Since the MIDI! basis set has been developed for the specific purpose of giving accurate geometries, rather than energies, it should reasonably be most useful in dual level or compound strategies for prediction of energies.

The HF/MIDI! method seems promising for use in nucleic acid chemistry and in particular for compounds containing phosphoryl functions. Since it gave good geometries for the phosphorus containing compounds of the original test set, it seemed valuable to extend the set to include more phosphorus containing structures. In particular, we wish to evaluate methods that could be useful in calculations carried out to clarify details of the mechanism of condensation in the H-phosphonate approach to oligonucleotides.[3–8] In addition, we felt that it would be valuable to assess in more detail how the different bond lengths and angles were affected, and above all how much the deviations in geometry affect the energies of the systems as obtained from subsequent single point calculations.

In this study we have compared the HF/MIDI![1] geometry optimizations on phosphorus acid derivatives with several other more widely used methods of relatively low computational cost. Deviations in bond angles and bond lengths are reported relative to not only the MP2/cc-pVDZ geometries,[9–11] as in the original evaluation of HF/MIDI!,[1] but also to the more widely used B3LYP/6-31G(d) geometry optimization method.[12–14] The impact on geometry differences on subsequent single point energy calculations was evaluated by using the popular B3LYP functional[12,13] with the 6-31+G(d,p) basis set. This method, together with B3LYP/6-31G(d) geometry optimization, has been extensively evaluated also for the G2 test set, and gives a mean absolute deviation of 4 kcal mol^{-1}.[14] Energies have also been calculated with the new EDF1 functional[15] that has been reported to give more accurate energies than B3LYP with the 6-31+G(d) basis set (MAD 1.7 kcal mol^{-1} for the G2 set with MP2 geometries). The molecules studied consisted of the HCOP test set as used by Truhlar and coworkers,[1] extended with eight more molecules, in order to cover a wider range of phosphorus compounds (Figure 1).

METHODS

The molecules and initial geometries were created with the GaussView interface to Gaussian98. The starting geometry for each molecule was then geometry optimized with the Gaussian98 program (rev. A.7)[16] and the default Berny algorithm using redundant internal coordinates.[17] For SCF convergence the default DIIS algorithm was used except in one case when quadratic convergence was used. The Möller-Plesset (MP2), Hartree-Fock, and B3LYP methods were used with the basis sets cc-pVDZ[9,10] (MP2), 6-31G(d),[18–21] (HF and B3LYP), 3-21G(d),[22–24] (HF), and MIDI![1] (HF and B3LYP). Inclusion of diffuse functions to the 6-31G(d) basis set only gave small differences in geometries in both HF and B3LYP calculations, with slightly higher deviations from the MP2 geometries (data not shown).

Subsequent B3LYP/6-31+G(d,p) single point energy calculations on the geometries from the respective optimizations were carried out using the default DIIS algorithm with the SCF = Tight option for the SCF convergence. The EDF1/6-31+G(d) single point energy calculations were performed using the Q-Chem program (version 1.2)[25] on the imported standard cartesian coordinates from the geometry optimizations in Gaussian98. The Gaussian98 calculations were performed on Compaq-Digital XP-1000 personal workstations with single 500 MHz EV6 processors and the Q-Chem calculations were performed either on

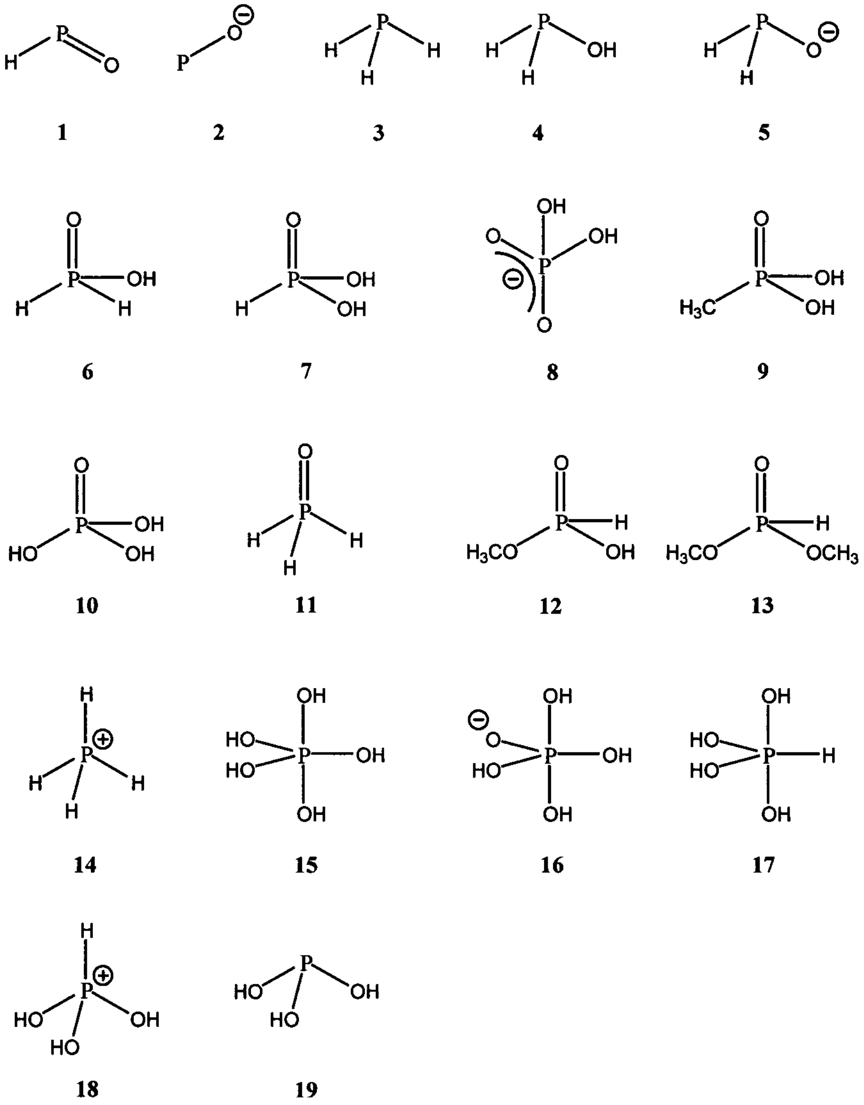


FIGURE 1 Model compounds used in the calculations.

the Compaq stations or using an SGI Octane (2 × 250 MHz, R10000) workstation.

For all geometry optimizations the deviations of all different types of bond angles and bond lengths, except those involving C—H bonds, have been calculated. These have then been compared to the

MP2/cc-pVDZ and B3LYP/6-31G(d) geometries and the deviations have been calculated for each type of bond and angle. For all geometry optimized structures the EDF1/6-31+G(d) and B3LYP/6-31+G(d,p) single point energies have been calculated. The deviations from the corresponding energies obtained from the MP2/cc-pVDZ and B3LYP/6-31G(d) geometries have been obtained, as well as the root mean square (RMS) and mean absolute deviations for the whole test set.

RESULTS AND DISCUSSION

The deviations of the obtained geometries from the references, in bond angles and bond lengths, are presented in Table I. Some trends are clearly seen. The HF geometries with the widely used 3-21G(d) and 6-31G(d) basis sets generally give the largest mean bond angle deviations (MBAD), when compared to the MP2/cc-pVDZ reference, most severely for H—O—P and C—O—P angles (between 5 and 11 degrees too large). The B3LYP methods give geometries much closer to the MP2 geometries, with the same angles being the most deviating ones. Worth noting is that the B3LYP functional in most cases give slightly better angles with the MIDI! basis set.

For the HF/MIDI! geometries all compared mean bond angle deviations are within 2 degrees, and two thirds of them within less than 1 degree of the MP2/cc-pVDZ geometries. In fact, out of all geometry optimization methods used, HF/MIDI! displays the lowest maximum deviation in bond angles from those of the MP2 geometries. The total picture is that HF/MIDI! and the B3LYP methods give quite MP2- like bond angles, while the HF/3-21G(d) and HF/6-31G(d) methods give considerably higher deviations from the MP2/cc-pVDZ geometries. The only slightly problematic bond angles using HF/MIDI! were in the conjugated anionic O—P—O system **8** and the pentacoordinated anion **16**. In these cases some of the H—O—P or O—P—O angles deviated more compared to the MP2 geometry (2–4 degrees). Some O—P—O and HO—P—OH angles were too small and O—P—OH angles too large in **8**, and some HO—P—OH too large and H—O—P to small in **16**. The other HF methods and B3LYP/MIDI! also gave deviations of the same order for the same angles in **8**. All methods gave similar or considerably higher (for the HF methods) deviations for the above angles in **16**. It also should be noted that with a starting geometry of **17** with all P—OH bonds equal, all HF methods converged to a local minimum having the P—H hydrogen in an apical position. However, a starting geometry with two P—OH bonds longer than the others converged (with all methods) to the lowest energy structure with an equatorial P—H hydrogen.

TABLE I Deviations of Bond Angles and Bond Lengths from Reference Geometries

Bond angle	MBAD cf MP2[a]	MBAD cf B3LYP[b]	Bond length	MBLD cf MP2 (Å)[c]	MBLD cf B3LYP (Å)[d]
HF/MIDI!					
H—O—P	−0.0705	−2.19	P—H	0.00520	0.00710
C—O—P	0.0393	−3.01	P=O	−0.00962	0.00886
O—P—O	0.235	0.0366	P—O	−0.0142	0.00219
H—P—O	0.489	0.170	O—H	−0.000601	−0.00243
H—P=O	−0.603	0.154	P—C	0.0243	0.0177
C—P—O	2.13	1.10	C—O	0.00713	0.00691
C—P=O	−0.586	0.475			
O—P=O	−1.63	−1.42			
H—P—H	1.21	1.14			
HF/3-21G(d)					
H—O—P	11.1	8.98	P—H	−0.0307	−0.0288
C—O—P	7.86	4.81	P=O	−0.0434	−0.0249
O—P—O	−0.0476	−0.246	P—O	−0.0574	−0.0410
H—P—O	0.624	0.305	O—H	−0.0046	−0.00644
H—P=O	0.0228	0.780	P—C	−0.0256	−0.0323
C—P—O	1.55	0.516	C—O	0.0196	0.0194
C—P=O	−1.00	0.0609			
O—P=O	−1.64	−1.43			
H—P—H	0.639	0.562			
HF/6-31G(d)					
H—O—P	4.98	2.86	P—H	−0.0289	−0.0270
C—O—P	5.21	2.16	P=O	−0.0486	−0.0301
O—P—O	0.202	0.0753	P—O	−0.0483	−0.0319
H—P—O	0.492	0.172	O—H	−0.0209	−0.0228
H—P=O	−0.986	−0.229	P—C	−0.00668	−0.0134
C—P—O	2.19	1.16	C—O	−0.0184	−0.0186
C—P=O	−1.78	−0.723			
O—P=O	−1.23	−1.02			
H—P—H	1.28	1.21			
B3LYP/MIDI![e]					
H—O—P	−3.10	−5.22	P—H	0.0280	0.0299
C—O—P	−1.47	−4.52	P=O	0.0195	0.0380
O—P—O	0.101	−0.0969	P—O	0.0114	0.0277
H—P—O	0.167	−0.153	O—H	0.0108	0.00892
H—P=O	0.0414	0.798	P—C	0.0360	0.0293
C—P—O	0.983	−0.0499	C—O	0.0251	0.0248
C—P=O	0.113	1.17			
O—P=O	−0.863	−0.654			
H—P—H	0.0364	−0.0410			

(Continued on next page)

TABLE I Deviations of Bond Angles and Bond Lengths from Reference Geometries *(Continued)*

Bond angle	MBAD cf MP2[a]	MBAD cf B3LYP[b]	Bond length	MBLD cf MP2 (Å)[c]	MBLD cf B3LYP (Å)[d]
B3LYP/6-31G(d)					
H—O—P	2.12		P—H	−0.00191	
C—O—P	3.05		P=O	−0.0185	
O—P—O	0.127		P—O	−0.0164	
H—P—O	0.319		O—H	0.00183	
H—P=O	−0.757		P—C	0.00669	
C—P—O	1.03		C—O	0.000220	
C—P=O	−1.06				
O—P=O	−0.208				
H—P—H	0.0774				

[a] Mean bond angle deviation from the MP2/cc-pVDZ geometries for test set molecules **1–19**.

[b] Mean bond angle deviation from the B3LYP/6-31G(d) geometries for test set molecules **1–19**.

[c] Mean bond length deviation from the MP2/cc-pVDZ geometries for test set molecules **1–19**.

[d] Mean bond length deviation from the B3LYP/6-31G(d) geometries for test set molecules **1–19**.

[e] Geometry optimization of molecule **2** with B3LYP/MIDI! was carried out using quadratically convergent SCF.

As with the bond angles, the mean bond length deviations (MBLD) relative to the MP2 geometries are most severe with HF/3-21G(d) and HF/6-31G(d), but these are now also joined by B3LYP/MIDI! (Table I). These HF calculations generally give too short bond lengths (with deviations of up to 0.05–0.06 Å) whereas the B3LYP/MIDI! method gives too long bonds (with deviations of up to 0.03–0.04 Å). The B3LYP/6-31G(d) calculations gave results that where quite similar to MP2/cc-pVDZ, with most bond lengths deviating only in the third decimal place, the exceptions being the P—O and P=O bonds that were 0.016–0.019 Å too short.

The HF/MIDI! calculations gave mean bond length deviations (MBLDs) relative to the MP2/cc-pVDZ geometries that were of a quality similar to those of the B3LYP/6-31G(d) calculations, giving a lower MBLD for P=O bonds and higher for P—C bonds. The only cases that gave significantly higher bond length deviations than the MBLDs were models **2** and **16**. The P—O bond length obtained for **2** was 0.064 Å shorter and the P—OH bond lengths obtained for **16** were 0.02–0.04 Å shorter than those of the MP2 geometries. For **2** the obtained bond lengths were even shorter for the other HF-calculations (0.076 and 0.094 Å) and for the B3LYP geometries they were also too short, although deviating less (0.018 Å with MIDI! and 0.045 Å with 6-31G(d)). As opposed to the bond angles, the bond length deviations in the phosphate monoanion **8** fall within the MBLD's reported in Table I. Overall the HF/MIDI! optimizations give geometries of near MP2 quality and at least as close to the MP2 geometries as those obtained with the more time demanding B3LYP/6-31G(d) optimizations.

Compared to the B3LYP/6-31G(d) geometries both the HF/MIDI! bond angles and bond lengths are in general closer than for any of the other methods, including B3LYP/MIDI! (Table I). The bond lengths vary only in the third or fourth decimal place, except for the P—C bonds. Hence the HF/MIDI! geometries can also be considered as B3LYP/6-31G(d)-like, and considerably more so than the other HF or the B3LYP/MIDI! geometries. Clearly the HF/MIDI! geometries deviate only slightly from the geometries obtained with the more costly MP2 and B3LYP methods. An important question is how much these small deviations affect the energies of the structures.

When comparing the outcome of subsequent single point energy calculations on the differently optimized geometries (Table II) it is clear that the MP2/cc-pVDZ and B3LYP 6-31G(d) geometries give very similar energies. The B3LYP/6-31+G(d,p) energies generally lie within one kcal mol^{-1} (MAD 0.76 kcal mol^{-1}) and the EDF1 energies within a few tenths of a kcal (MAD 0.26 kcal mol^{-1}) of each other. For most purposes these geometry optimization methods can be considered to provide geometries that give equivalent energies for these types of molecules. From Table II it also is clear that the HF/3-21G(d), HF/6-31G(d), and B3LYP/MIDI! geometries give consistently too high energies. For these Hartree-Fock methods the deviation is worst when EDF1 energies are compared (up to +5–8 kcal mol^{-1} with a MAD of 3–4 kcal mol^{-1}) and for the B3LYP/MIDI! geometries the B3LYP/6-31+G(d,p) energies deviate most (up to 4–5 kcal mol^{-1} with a MAD of 2.27 kcal mol^{-1}).

Upon subsequent single point energy calculation, the HF/MIDI! geometry optimized structures give energies that are quite comparable to the ones obtained with the MP2/cc-pVDZ geometries. Both the B3LYP/6-31+G(d,p) and EDF1/6-31+G(d) energies are usually within one kcal mol^{-1} (MAD 0.55 and 0.50 kcal mol^{-1}) and in many cases within a few tenths of a kcal mol^{-1}. The three models that with HF/MIDI! geometry give the highest deviations in single point energies from the energies of the MP2 geometries are the anions **2**, **8**, and **16**. These were also the models that gave the HF/MIDI! geometries that deviated most from the reference. The energies for **8**, nevertheless, deviate less than for the geometries obtained with the other HF basis set combinations and B3LYP/MIDI!, but it indicates that caution

TABLE II Deviation of Single Point Energies for Different Geometry Optimization Methods from Single Point Energies of MP2/cc-pVDZ

	ΔE (kcal/mol) HF/MIDI! geometry opt		ΔE (kcal/mol) HF/3-21G(d) geometry opt		ΔE (kcal/mol) HF/6-31G(d) geometry opt		ΔE (kcal/mol) B3LYP/MIDI! geometry opt		ΔE (kcal/mol) B3LYP/6-31G(d)	
Molecule	**A** B3LYP/ 6-31+G(d,p)	**B** EDF1/ 6-31+G(d)	**A** B3LYP/ 6-31+G(d,p)	**B** EDF1/ 6-31+G(d)	**A** B3LYP/ 6-31+G(d,p)	**B** EDF1/ 6-31+G(d)	**A** B3LYP/ 6-31+G(d,p)	**B** EDF1/ 6-31+G(d)	**A** B3LYP/ 6-31+G(d,p)	**B** EDF1/ 6-31+G(d)
1	−0.56	0.01	0.56	1.68	1.60	2.31	0.31	0.11	−0.54	−0.04
2	−1.65	[a]	−1.57	[a]	0.45	[a]	−0.76	[a]	−1.50	[a]
3	0.10	0.10	0.39	[b]	0.31	0.65	0.38	0.15	0.00	−0.02
4	0.05	0.11	0.81	1.45	0.66	1.38	0.99	0.48	−0.15	−0.04
5	−0.17	0.03	0.89	1.65	1.08	1.76	1.01	0.52	0.13	0.26
6	0.15	0.54	1.81	3.55	1.48	3.37	2.08	1.13	−0.42	0.04
7	−0.12	0.41	2.50	5.06	1.67	3.99	2.83	1.62	−0.72	0.00
8	1.62	2.62	2.84	5.72	2.29	5.15	4.32	3.22	−0.57	0.34
9	−0.22	0.37	2.65	5.27	1.69	3.83	2.28	1.15	−0.71	0.00
10	−0.36	0.30	3.28	6.68	2.04	4.83	3.57	1.94	−0.95	0.09
11	−0.10	−0.04	0.91	1.72	1.28	2.09	1.36	0.79	−0.15	0.08
12	−0.26	0.46	2.34	4.79	1.79	3.88	2.42	1.43	−0.90	−0.12
13	−0.95	0.57	2.71	5.28	2.61	3.99	2.22	1.54	−1.02	−0.24
14	0.06	0.02	0.52	0.74	0.45	0.74	0.54	−3.07	−0.01	0.02
15	−0.15	0.59	3.15	6.38	2.07	5.27	4.49	2.05	−0.85	−0.31
16	−3.21	−1.71	4.75	8.25	−0.66	2.99	4.37	2.51	−3.74	−2.84
17	0.04	0.44	2.62	5.01	1.70	4.12	3.76	1.97	−0.65	−0.21
18	0.68	0.58	3.72	6.75	0.77	3.08	4.71	2.86	−0.91	−0.11
19	0.06	0.52	1.74	3.58	1.18	3.00	0.06	0.52	−0.59	−0.21
MAD	0.55	0.50	2.09	3.87	1.36	2.97	2.27	1.43	0.76	0.26
RMS	0.95	0.78	1.48	2.22	0.80	1.41	1.74	1.39	0.83	0.68

[a] Did not converge.
[b] Calculation failed.

TABLE III Deviation of Single Point Energies for Different Geometry Optimization Methods from Single Point Energies of B3LYP/6-31G(d) Geometries

	ΔE (kcal/mol) HF/MIDI! geometry opt		ΔE (kcal/mol) HF/3-21G(d) geometry opt		ΔE (kcal/mol) HF/6-31G(d) geometry opt		ΔE (kcal/mol) B3LYP/MIDI! geometry opt	
Molecule	**A** B3LYP/ 6-31+G(d,p)	**B** EDF1/ 6-31+G(d)	**A** B3LYP/ 6-31+G(d,p)	**B** EDF1/ 6-31+G(d)	**A** B3LYP/ 6-31+G(d,p)	**B** EDF1/ 6-31+G(d)	**A** B3LYP/ 6-31+G(d,p)	**B** EDF1/ 6-31+G(d)
1	−0.01	0.04	1.10	1.72	1.59	2.35	0.85	0.14
2	−0.15	[a]	−0.08	[a]	0.30	[a]	0.74	[a]
3	0.10	0.12	0.38	[b]	0.41	0.66	0.38	0.17
4	0.19	0.15	0.96	1.49	0.85	1.42	1.14	0.52
5	−0.30	−0.23	0.76	1.38	0.77	1.49	0.88	0.26
6	0.57	0.51	2.23	3.52	2.05	3.34	2.50	1.09
7	0.60	0.40	3.22	5.06	2.27	3.98	3.55	1.62
8	2.19	2.28	3.41	5.37	2.85	4.81	4.89	2.87
9	0.49	0.37	3.36	5.27	2.18	3.83	2.99	1.15
10	0.59	0.21	4.24	6.59	2.63	4.73	4.52	1.85
11	0.05	−0.12	1.06	1.64	1.32	2.01	1.51	0.71
12	0.65	0.58	3.24	4.91	2.43	4.00	3.32	1.55
13	0.08	0.81	3.73	5.51	2.68	4.22	3.25	1.78
14	0.07	−0.01	0.53	0.72	0.46	0.72	−0.76	−2.52
15	0.70	0.90	4.00	6.69	2.92	5.58	5.34	2.36
16	0.53	1.13	8.49	11.09	3.08	5.83	8.11	5.35
17	0.69	0.64	3.27	5.21	2.35	4.32	4.41	2.17
18	1.59	0.69	4.63	6.85	1.68	3.19	5.62	2.96
19	0.66	0.73	2.33	3.80	1.78	3.21	0.66	0.73
MAD	0.54	0.52	2.69	4.04	1.82	3.14	2.92	1.57
RMS	0.59	0.58	2.03	2.64	0.91	1.57	2.27	1.60

[a] Did not converge.

[b] Calculation failed.

is in place and one probably should count on larger errors when dealing with compounds having this type of functionality. The energies for models **2** and **16** are deviating from the references substantially for most methods, but more inconsistently. In contrast to the other models, **16** gives the B3LYP/6-31G+(d,p) energy closest to the reference with the HF/6-31G(d) geometry, while the EDF1 energy is closest to the reference using the HF/MIDI! geometry. For the methods that give the best overall performance, HF/MIDI! and B3LYP/6-31G(d), the energies for **2** are similar (-1.65 and -1.5 kcal mol^{-1} respectively) and the energies for **16** closest to the reference are those for HF/MIDI! (-1.7 to -3.2 kcal mol^{-1}) rather than those for B3LYP/6-31G(d) (-2.8 to -3.7 kcal mol^{-1}).

When comparing the single point energies of the differently optimized models with B3LYP/6-31G(d) as reference, not unexpectedly, a similar picture as for the geometries emerges (Table III). The HF/3-21G(d), HF/6-31G(d), and B3LYP/MIDI! geometries give energies that deviate substantially from those of the B3LYP/6-31G(d) geometries (up to 5–11 kcal mol^{-1} with MADs of 1.5 to 4 kcal mol^{-1}). The HF/MIDI! geometries, however, give energies generally within one kcal mol^{-1} (MAD about 0.5 kcal mol^{-1}) of those of the B3LYP/6-31G(d) geometries, with the only exceptions being models **8** and **18** that have slightly higher deviations (1.6–2.3 kcal mol^{-1}).

Using HF/MIDI! for geometry optimization of phosphorus acids clearly gives geometries and subsequent single point energies quite comparable to those of considerably more computationally demanding methods. To obtain a crude estimate of the difference in computational cost we have compared the time required for the different geometry optimizations. The comparison is quite approximate since the number of optimization cycles can vary substantially and we only have correlated the time with number of heavy atoms, which of course gives a weighted contribution from the heavier phosphorus atom. It should anyway be useful as a rough guide as to the limits of the methods that give quite comparable geometries and energies.

As indicated in Figure 2 there is a tremendous difference in the dependence of CPU time on the size of the molecule, for geometry optimizations using the best of the different methods. To obtain an estimate on the limits of each method we have made a rough approximation by correlating the CPU time to the number of heavy atoms via a simple exponential equation (i.e., CPU time = (# heavy atoms)^{X}). We have, using the equations of the curve fits of Figure 2, obtained estimates of expected computational times for differently sized molecules (Table IV). These correspond to examples of interest in nucleic acid chemistry; typical examples could be isopropyl pivaloyl H-phosphonate, a model of an

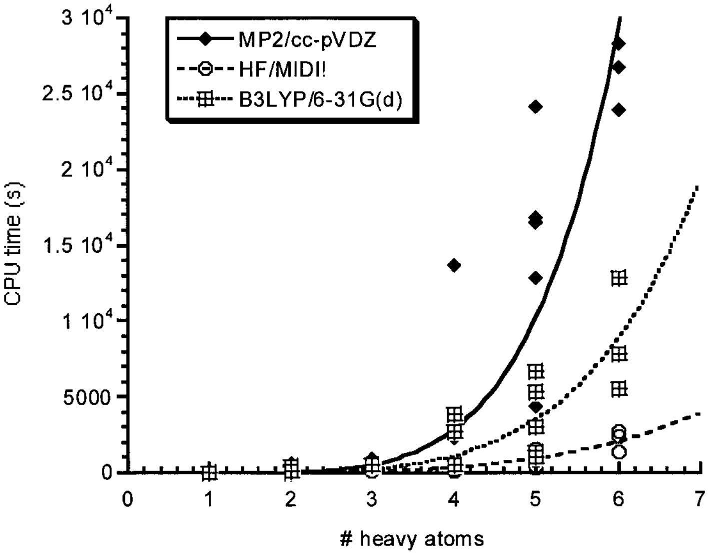


FIGURE 2 CPU time for geometry optimization using different methods versus number of heavy atoms. CPU time = (# heavy atoms)^{X}; $X_{(MP2/cc\text{-}pVDZ)}$: 5.74, $X_{(B3LYP/6\text{-}31G(d))}$: 5.07, $X_{(HF/MIDI!)}$: 4.26.

intermediate in H-phosphonate coupling[26] (13 heavy atoms), a purine nucleoside (20 heavy atoms), and a 2′-O-alkyl nucleoside 3′-(ethyl phosphate) (ca 30 heavy atoms).

Already the largest models in our test set show a considerable difference in computational cost. Upon increasing the number of heavy atoms to 13, geometry optimization with HF/MIDI! is still quite painless while the least costly of the other methods would take some days and MP2 calculations would be quite time-demanding. For even the larger molecules B3LYP is inconveniently slow and MP2 is hardly reasonable even with much more processing power, especially considering the small increase in accuracy that would be achieved relative to using HF/MIDI!.

TABLE IV Approximate Expected CPU Times for Geometry Optimization of Differently Sized Molecules with the Three Most Accurate Methods

# Heavy atoms	CPU time MP2/cc-pVDZ	CPU time B3LYP/6-31G(d)	CPU time HF/MIDI!
6	8 h	2.5 h	35 min
13	1 month	>5 days	16 h
20	1 year	1.5 months	4 days
30	10 years	1 year	24 days

CONCLUSIONS

Geometry optimizations of phosphorus acid derivatives using HF/MIDI! are computationally inexpensive and give geometries that are close to those obtained with the MP2/cc-pVDZ as well as B3LYP/6-31G(d), and superior to those obtained with HF/3-21G(d) and HF/6-31G(d). Most importantly, the HF/MIDI! geometries lead to B3LYP/6-31+G(d,p) and EDF1/6-31+G(d) single point energies of a mean absolute deviation (MAD) of 0.5–0.55 kcal mol^{-1} from those obtained using MP2/cc-pVDZ geometries, which is comparable to results with B3LYP/6-31G(d) geometries (MAD (from MP2) of 0.3–0.8 kcal mol^{-1}). This deviation is quite small for many practical applications and, especially if computational cost is essential, the method is clearly of choice for molecules with more than 10 heavy atoms. Since the combination B3LYP/6-31+G(d,p)//B3LYP/6-31G(d) gives energies with a MAD of 4 kcal mol^{-1} relative to experimental data[14] one can assume that B3LYP/6-31+G(d,p)//HF/MIDI! would not give a MAD of more than about 4–5 kcal mol^{-1}. The combination EDF1/6-31+G(d)//HF/MIDI! probably gives even more accurate energies compared to experimental data but to get a better estimate of this one would need more extensive comparisons.

In summary, for quantum chemical calculations of phosphorus acid derivatives use of the modest Hartree Fock level in combination with the MIDI! basis set gives near-MP2 quality in geometries, resulting in a fair accuracy of energies in subsequent single point calculations, at a low computational cost. The other methods that give similar accuracies require a much higher computational cost.

REFERENCES

[1] R. E. Easton, D. J. Giesen, A. Welch, C. J. Cramer, and D. G. Truhlar, *Theor. Chim. Acta*, **93,** 281 (1996).

[2] J. W. Ochterski, G. A. Petersson, and J. A. Montgomery, Jr., *J. Chem. Phys.*, **104,** 2598 (1996).

[3] P. J. Garegg, T. Regberg, J. Stawinski, and R. Strömberg, *Chemica Scripta*, **25,** 28 (1985).

[4] P. J. Garegg, T. Regberg, J. Stawinski, and R. Strömberg, *Chemica Scripta*, **26,** 5 (1986).

[5] P. J. Garegg, I. Lindh, T. Regberg, J. Stawinski, R. Strömberg, and C. Henrichson, *Tetrahedron Lett.*, **27,** 4051 (1986).

[6] P. J. Garegg, I. Lindh, T. Regberg, J. Stawinski, C. Henrichson, and R. Strömberg, *Tetrahedron Lett.*, **27,** 4055 (1986).

[7] B. C. Froehler and M. D. Matteucci, *Tetrahedron Lett.*, **27,** 469 (1986).

[8] B. C. Froehler, P. G. Ng, and M. D. Matteucci, *Nucl. Acids Res.*, **14,** 5399 (1986).

[9] T. H. Dunning Jr., *J. Chem. Phys.*, **90,** 1007 (1989).

[10] D. E Woon and T. H. Dunning Jr., *J. Chem. Phys.*, **98,** 1358 (1993).

[11] K. A. Peterson and T. H. Dunning Jr., *J. Chem. Phys.*, **106**(10), 4119 (1997).

[12] A. D. Becke, *J. Chem. Phys.*, **98,** 5648 (1993).

[13] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, *J. Phys. Chem.*, **98,** 11623 (1994).

[14] J. B. Foresman and Æleen Frisch, *Exploring Chemistry with Electronic Structure Methods* (Gaussian, Inc., Pittsburgh, PA, xxxx), 2nd ed., p. 157.

[15] R. D. Adamson, P. M. W. Gill, and J. A. Pople, *Chem. Phys. Lett.*, **284**, 6 (1998).

[16] M. J. Frisch, G. W. Trucks, H. B. Schlegel et al., Gaussian 98 (Rev. A.7), Gaussian, Inc., Pittsburgh, PA (1998).

[17] C. Peng, P. Y. Ayala, H. B. Schlegel, and M. J. Frisch, *J. Comp. Chem.*, **17,** 49 (1996).

[18] W. J. Hehre, R. Ditchfield, and J. A. Pople, *J. Chem. Phys.*, **56,** 2257 (1972).

[19] P. C. Hariharan and J. A. Pople, *Theo. Chim. Acta*, **28,** 213 (1973).

[20] M. S. Gordon, *Chem. Phys. Lett.*, **76,** 163 (1980).

[21] M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees, and J. A. Pople, *J. Chem. Phys.*, **90,** 3654 (1982).

[22] J. S. Binkley, J. A. Pople, and W. J. Hehre, *J. Amer. Chem. Soc.*, **102,** 939 (1980).

[23] W. J. Pietro, M. M. Francl, W. J. Hehre, D. J. Defrees, J. A. Pople, and J. S. Binkley, *J. Amer. Chem. Soc.*, **104,** 5039 (1982).

[24] M. S. Gordon, J. S. Binkley, J. A. Pople, W. J. Pietro, and W. J. Hehre, *J. Amer. Chem. Soc.*, **104,** 2797 (1982).

[25] C. A. White, J. Kong, D. R. Maurice et al., Q-Chem, Version 1.2, Q-Chem, Inc., Pittsburgh, PA (1998).

[26] P. J. Garegg, T. Regberg, J. Stawinski, and R. Strömberg, *Nucleosides and Nucleotides*, **6**(1/2), 283 (1987).